



FAKULTÄT FÜR
INFORMATIK

**Seminar: Datenqualität und Optimierungs- und Modellierungstechniken für
Datenbanken**

Workload bei Datenqualitätsaufgaben

17.12.2009

**Benjamin Röhl
benjamin.roehl@st.ovgu.de**

Agenda

- Begriffe
- Motivation
- Ansatz zur verteilten Bewertungsmethode
 - Aufbau der Architektur
 - Allokation von Queries
 - Ausführung von Queries
 - Aufbau des LQA
 - Phantom States
- Zusammenfassung
- Quellen

Begriffe

- Workload
- Datenqualitätsaufgaben

Begriffe: Workload

- Anfrage an Datenbank
- Einzelne Anfrage (SQL-Statement) oder
- Summe von Anfragen für eine Applikation, Benutzer, etc. in einem Zeitrahmen sein
- Antwortzeit reduzieren → Durchsatz messen

Begriffe: Datenqualitätsaufgaben

- Bewertung der Qualität von Daten
- Finden von ungenauen, nicht kompletten, nicht konsistenten oder nicht aktuellen Daten
- Korrigieren → Verbessern der Datenqualität

Motivation

- Eingeschränkte Ressourcen
- Geschäftsprozesse nicht stören oder verzögern
- Verteilte Datenbanken

→ Wann? Wie? Was?

- Verwaltung von Workloads

Ansatz zur verteilten Bewertungsmethode

Lokale Qualitäts-Assessor

- Wiederholende Ausführung
- Entsprechend Workload-Muster
- Transfer von Zwischen- und Endergebnissen

Globale Qualitäts-Assessor

- Verteilt Queries zu lokalen Seiten
- Entsprechend Kostenmodell
- Verwaltet Depot von Endergebnissen

Aufbau der Architektur

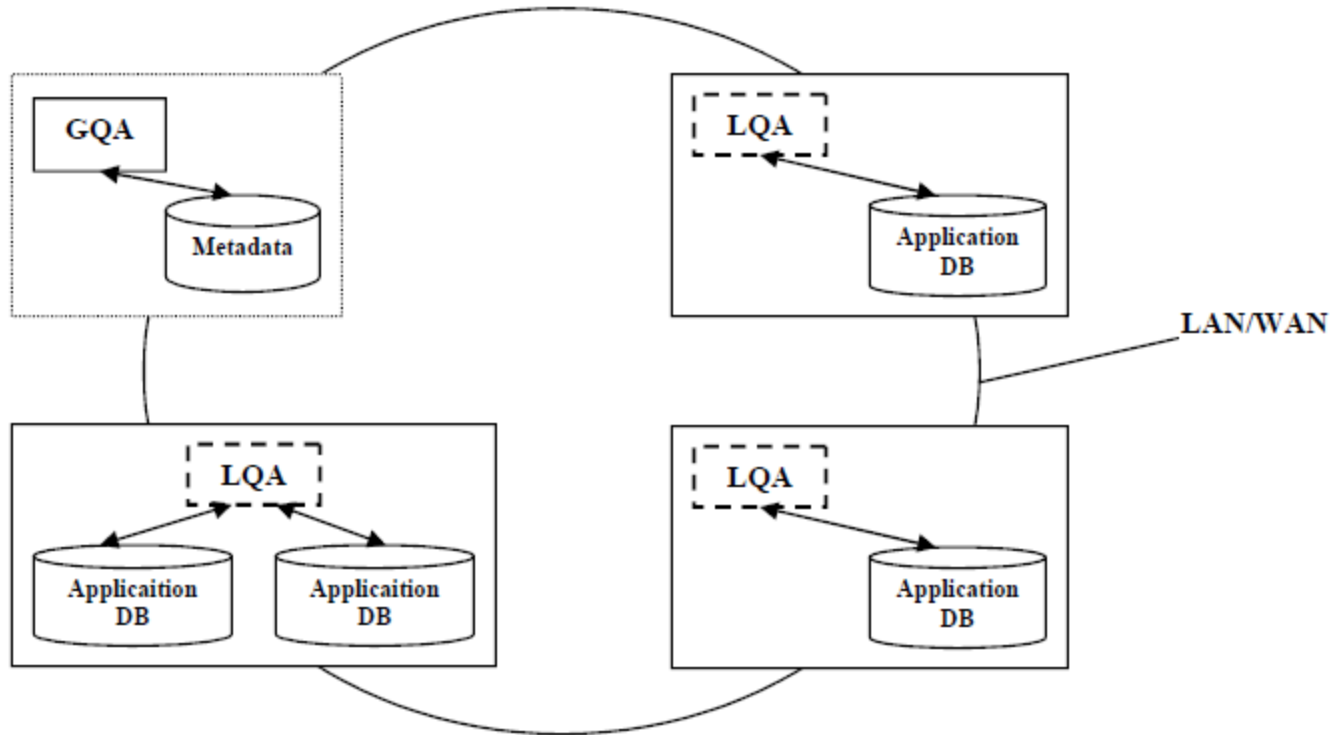
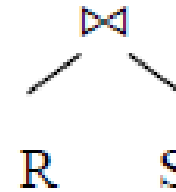


Abb.1: System Architektur, *Non-Intrusive Assessment of Organisational Data Quality*, 2001

Allokation von Queries

- Transformation der Queries
 - Referenzieren lokaler Tabellen und Attribute
 - Fragmentierung in Sub-Queries
 - Viele Zustellmöglichkeiten
 - z.B. Verbund 2er Tabellen
- Minimierung der Transferzeit



Ausführung von Queries

- Lokale Seite bestimmt Workload-Muster
- Vorhersehung von Busy- und Idle Phasen

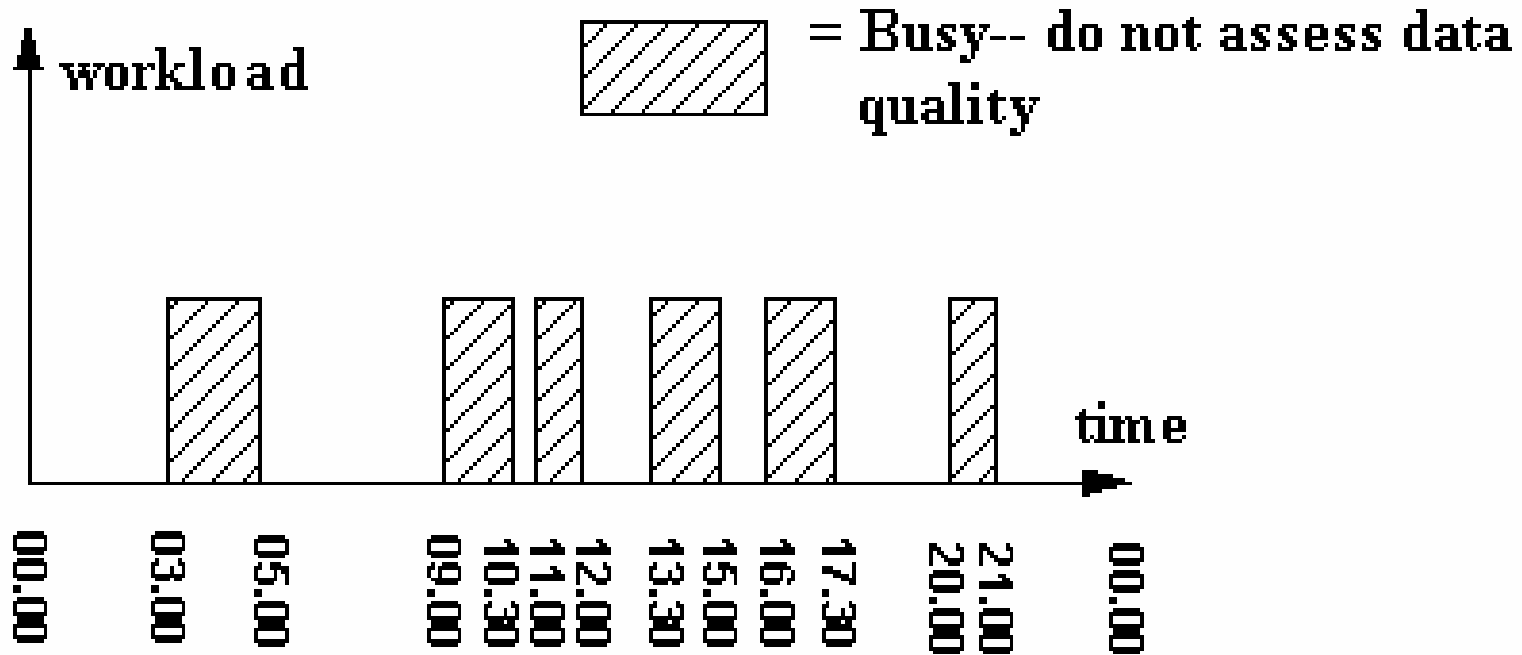


Abb.2: Beispiel für Zeitplan von Workloads,
Non-Intrusive Assessment of Organisational Data Quality, 2001

Aufbau des LQA

- Operation Metadata – enthält Sub-Queries
- Intermediate Data – enthält Zwischenergebnisse von LQAs
- Query Evaluator – Ausführung der Sub-Queries

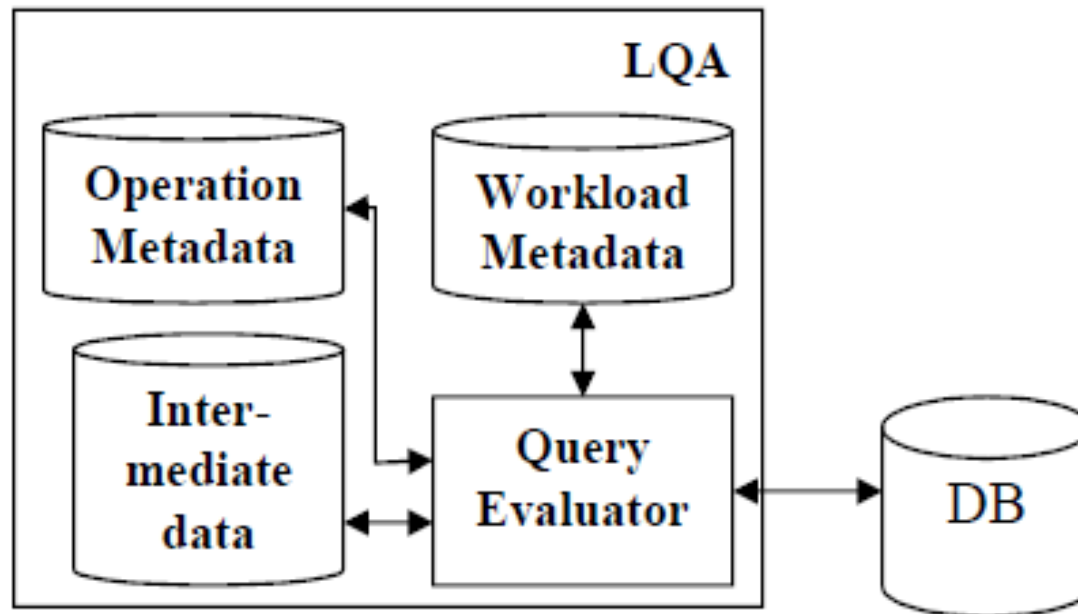


Abb.3: Lokale Qualitäts-Assessor Architektur,
Non-Intrusive Assessment of Organisational Data Quality, 2001

Phantom States

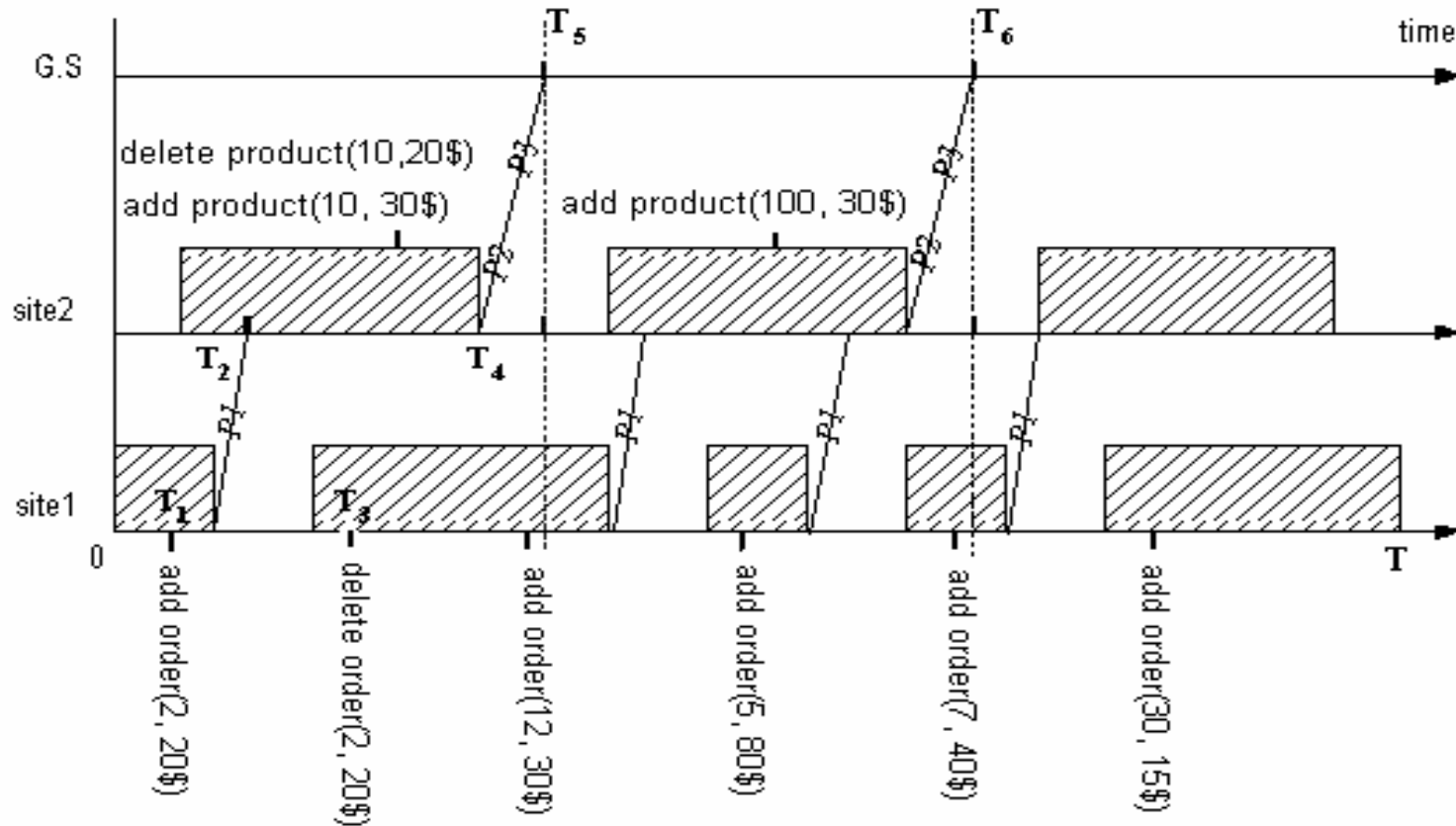


Abb.4: Beispiel zur Bewertungsausführung,
Non-Intrusive Assessment of Organisational Data Quality, 2001

→ **Kostenmodell**

Zusammenfassung

- Häufiger ausführbar
- Ohne Geschäftsprozesse zu stören
- Nicht akkurate Darstellung der Datenqualität
- Verzögerung bei verteilter Query-Ausführung

Quellen

- J. F. Koopmann: **Investigating the workload in database systems to improve throughput**, 03.03.2009
- C. K. Gupta, A. Mehta, U. Dayal: **Scheduling Memory Usage Of A Workload**, 09.07.2009
- B. Jin, S. M. Embury, **Non-Intrusive Assessment of Organisational Data Quality**, 2001

Vielen Dank für Ihre Aufmerksamkeit!

www.ovgu.de